

Self-supervised pretraining for transferable quantitative phase image cell segmentation

TOMAS VICAR,^{1,2}  JIRI CHMELIK,¹  ROMAN JAKUBICEK,¹ LARISA CHMELIKOVA,¹ JAROMIR GUMULEC,²  JAN BALVAN,²  IVO PROVAZNIK,¹ AND RADIM KOLAR^{1,*} 

¹Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

²Department of Pathological Physiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
*kolarr@vutbr.cz

Abstract: In this paper, a novel U-Net-based method for robust adherent cell segmentation for quantitative phase microscopy image is designed and optimised. We designed and evaluated four specific post-processing pipelines. To increase the transferability to different cell types, non-deep learning transfer with adjustable parameters is used in the post-processing step. Additionally, we proposed a self-supervised pretraining technique using nonlabelled data, which is trained to reconstruct multiple image distortions and improved the segmentation performance from 0.67 to 0.70 of object-wise intersection over union. Moreover, we publish a new dataset of manually labelled images suitable for this task together with the unlabelled data for self-supervised pretraining.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Quantitative phase imaging (QPI) has proved to be a powerful tool for label-free live cell microscopy. This technique typically provides images with superior image properties with respect to automated image processing [1]. Various QPI techniques have been developed and tested during the last decades, utilising different setups, e.g., off-axis, in-line or phase-shifting [2]. Ongoing progress in QPI microscopy enables the time-lapse observation of subtle changes in the quantitative phase dynamics of cells, such as cell dry mass distribution. It has been shown (e.g. [3,4]) that QPI-measured dynamical changes of various parameters are typical for specific cell behaviour and can be used in different applications, e.g., cell motility assessment, homogeneity of cell content or cell mass distribution evaluation. These phase-related changes can be observed without fixation, labelling, or cell harvesting, which might severely change cell characteristics [4].

A large body of recently published papers show that instance segmentation is still critical for microscopy image segmentation in general, and QPI needs its specific setting. As we have shown [1], the cell instance segmentation based on QPI image data typically provides better results in comparison to other microscopic imaging techniques (e.g., phase contrast, differential interference contrast etc.), and relatively basic image processing techniques can provide sufficient results. However, there are still applications where precise cell segmentation is crucial due to morphological parameters derived from individually segmented cells. This is particularly important in cell death detection [4], cell cycle detection [5] or quantification of cell culture quality [6] in a label-free setup utilizing QPI. Basic image processing methods can perform well in many cases, as shown by a combination of thresholding, hole filling, and watershed methods for yeast cell segmentation [7], Otsu-based thresholding of murine melanoma cells [8], thresholding and watershed algorithms for adherent/suspended cell classification [9], iterative thresholding method [10], or improved iterative thresholding using Laplacian of Gaussian image enhancement and distance transform-based splitting for dense cell clusters [11].

Fully convolutional neural networks (CNN), e.g., U-Net [12], with specific modifications for individual cell separation, can be successfully applied in these applications. However, direct application of U-Net for binary segmentation (foreground-background) does not achieve robust separation of individual cells because each error in boundary pixels results in the connection of these cells into one segmented cell. This can be overcome by a suitable modification of the network output, as demonstrated on various microscopic non-QPI image data [13–16]. One possibility is to introduce three classes, where a 'thicker cell boundary' class is introduced, and, after prediction, this boundary is used to divide the cells into individual objects [13,14]. Another simple solution is a prediction of a distance transform of the cell segmentation mask, where the foreground can be found by thresholding, and individual cells in the prediction can be found with the maxima detector. Another approach predicts the distance to a neighbouring cell or combines these multiple approaches together [15]. There is an even more complex solution for the prediction of star convex polygons, where for each pixel, distance to the boundary in several directions is predicted – StarDist [16], or prediction of a vector field that can be used for cell separation – CellPose [17]. Furthermore, specific deep learning approaches have also been proposed for complex cell analysis of QPI data. Mask region-based convolution neural (Mask R-CNN [18]) network was used in two recent papers [19] [20]. A U-Net architecture [12] was also applied to QPI images, for instance, Yi et al. [21] applied U-Net to red blood cell segmentation directly on hologram images to avoid the image reconstruction part. A similar method using QPI of adherent mesenchymal stem cells was reported by Zhang et al. [22].

Recently, self-supervised pretraining methods became a popular and successful way to improve the performance of deep learning (DL) methods [23]. Currently the best performing methods based on contrastive learning (SimCLR [20] and SimCLRv2 [19]) are not suitable for segmentation tasks, because they require network architecture with classification output, thus, they cannot be used for pretraining of segmentation architectures (it can be used only for pretraining of encoder part of the network). Several other approaches for self-supervised pretraining have shown promising results, including prediction of image rotation [24], solving a jigsaw puzzle [25], image in-painting prediction [26] or denoising [27], where only the last two are suitable for segmentation networks. SeSe-Net [28] propose a more complex self-supervised approach, where two networks are trained; one is trained for the segmentation quality prediction and another for the segmentation. These two networks can then be applied for training on unlabelled data.

In this paper, we have implemented and compared four U-Net [12] based approaches for instance cell segmentation with four specifically designed post-processing pipelines using different image processing methods. To enable the transferability of the segmentation network to different sample types (i.e. different cell morphologies) without the need of annotated training data, we aimed to design these post-processing pipelines with only a few tunable parameters, which enables to perform a *non-deep learning transfer (non-DL transfer)*. Compared to standard transfer learning, this approach does not require training data and computational demanding training of DL model. We also aimed at the application of specific pretraining strategies using non-labelled images, which can be used for self-supervised pretraining to improve final segmentation quality. The proposed methodology with self-supervised pretraining improved both segmentation performance and transferability to different cell types. Moreover, we propose a new dataset suitable for this task. Besides manually labelled data, this dataset contains unlabelled data, which can be used for self-supervised pretraining.

In summary, our main contributions are:

- Four strategies for instance cell segmentation with U-Net are compared.
- Specific post-processing pipelines with tunable/optimizable parameters are designed for each segmentation strategy.

- Transferability to different cell types by optimisation of post-processing parameters are tested.
- The self-supervised pretraining method improving both the segmentation performance and transferability to different cell types is proposed.
- A new manually labelled quantitative phase imaging dataset for cell segmentation with unlabelled data for self-supervised pretraining is created.

2. Material and methods

2.1. Dataset

A set of adherent cell lines of various origins, tumorigenic potential, and morphology were used in this paper (PC-3, PNT1A, 22Rv1, DU145, LNCaP, A2058, A2780, Fadu, G361, HOB). PC-3, PNT1A, 22Rv1, DU145, LNCaP, A2780, and G361 cell lines were cultured in RPMI-1640 medium, A2058, FaDu, and HOB cell lines were cultured in DMEM-F12 medium, all supplemented with antibiotics (penicillin 100 U/ml and streptomycin 0.1 mg/ml), and with 10% fetal bovine serum (FBS). Prior to microscopy acquisition, the cells were maintained at 37 °C in a humidified (60%) incubator with 5% CO₂ (Sanyo, Japan). For acquisition purposes, the cells were cultivated in the Flow chamber μ -Slide I Luer Family (Ibidi, Martinsried, Germany). To maintain standard cultivation conditions during time-lapse experiments, cells were placed in the gas chamber H201 – for Mad City Labs Z100/Z500 piezo Z-stage (Okolab, Ottaviano NA, Italy). For the acquisition of QPI, a coherence-controlled holographic microscope (Telight, Q-Phase) was used. Objective Nikon Plan 10 \times /0.3 was used for hologram acquisition with a CCD camera (XIMEA MR4021MC). Holographic data were numerically reconstructed with the Fourier transform method (described in [29]) and phase unwrapping was used on the phase image. QPI datasets used in this paper were acquired during various experimental setups and treatments. The individual images were taken at different time intervals (at least three hours) during which the cells significantly changed their morphology. Thus we obtained morphologically distinct cells in all images of our dataset.

Our datasets consist of 244 labelled images of PC-3 (7,907 cells), 205 labelled PNT1A (9,288 cells) denoted as *QPI_Seg_PNT1A_PC3*, and 1,819 unlabelled images with a mixture of 22Rv1, A2058, A2780, DU145, Fadu, G361, HOB and LNCaP used for pretraining denoted as *QPI_Cell_unlabelled*. Data were labelled using a custom MATLAB semiautomatic tool, where the image is pre-segmented using [10], and then manually edited with a set of drawing tools (e.g., cell splitting with scribble, union of selected cells, drawing a new cell, deleting the selected cell and correction of cell borders). An example of PC-3 and PNT1A cells from *QPI_Seg_PNT1A_PC3* dataset is shown in Fig. 4. An example of images of another cell lines from *QPI_Cell_unlabelled* dataset is shown Supplement 1, Fig. S2. Dataset is available at the Zenodo repository [39] and the source code for semiautomatic segmentation is available together with all proposed algorithms at [40].

Labelled data were divided into training, validation, and testing sets in proportion 85/5/10% and pretraining data were divided into training and validation sets in portion 95/5%. Labelled data (training part) were also used for pretraining to make the pretraining set even larger.

2.2. Segmentation approaches

In this work, a novel approach for instance segmentation, inspired by [15] was designed and tested. Specifically, besides binary foreground segmentation, four other parametric images were predicted and used for splitting the foreground into individual cells. Specific post-processing (cell detection) pipelines to achieve instance cell segmentation were designed for each of these prediction approaches. A general processing scheme is shown in Fig. 1(a).

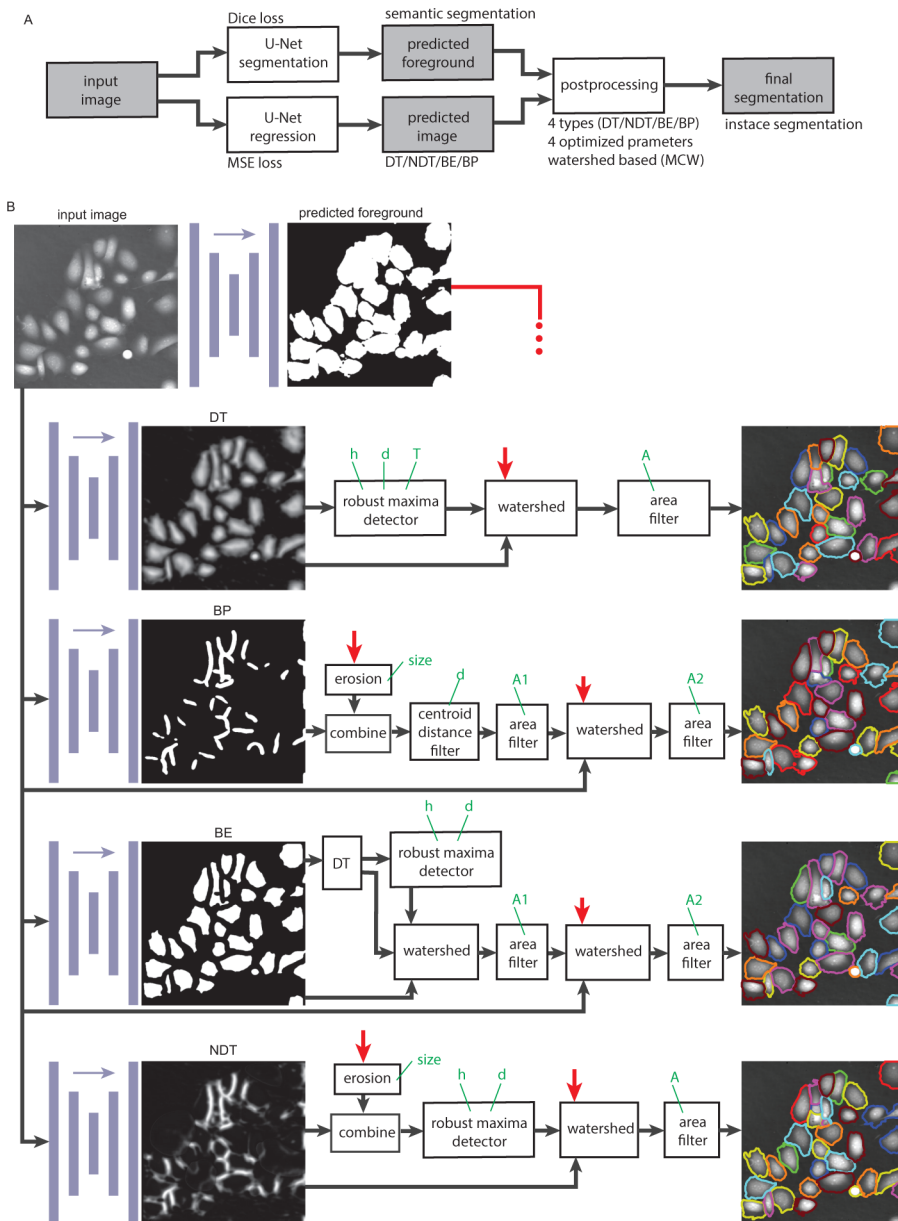


Fig. 1. Block diagrams of tested instance segmentation methods: (a) General processing schema. (b) Detailed processing scheme of individual post-processing methods (distance transform – *DT*, prediction of boundary pixels – *BP*, prediction of eroded image – *BE*, neighbour distance transform – *NDT*). Optimised parameters of the individual post-processing methods are shown in green. The red arrows indicate an input of the predicted foreground. U-Net prediction model is in blue colour.

The U-Net [12] network with EfficientNet-B2 [30] encoder (E-U-Net) was used in our approach. Two different loss functions were used. Dice loss was used for U-Net training of pixel classification outputs and Mean squared error (MSE) was used for the training of U-Net with pixel regression outputs. For more details about the implementation see [Supplement 1](#).

All proposed post-processing utilise marker-controlled watershed (MCW) [31] (similarly to [14]), which is a highly efficient method in cell segmentation tasks [1] using QPI data. These post-processing approaches were further extended by subsequent steps (see Fig. 1(b)) to make it efficient and easy to optimise by adjustment of four parameters using Bayesian optimisation [32]. In our implementation, the MCW has three inputs – (1) a binary mask (foreground mask), which is split into individual cells; (2) seeds, where every seed produce one object; (3) input image, which is used to generate the watershed borders (i.e., flooded image). Borders produced by the watershed are used to split the binary foreground mask. The last step of all methods is the area filter, which removes objects smaller than an optimized threshold. In the three post-processing pipelines, a robust maxima detector is used – it is a local maxima detector applying a constraint of the minimal distance d on the individual detected maxima; h -maxima transform [33] for a constraint of the minimal peak prominence; and threshold for minimal maxima value T .

Parameters of the post-processing pipeline were determined by Bayesian optimisation [32] (implementation from [34]), where the value of the cost function (Object-wise Intersection over Union, OIoU – see Section 2.4) was optimised on the validation set. The ranges of optimised parameters and the optimised values are summarized in Supplemental document. A brief description of the implemented and tested pipelines follows.

DT – In the first approach, a normalised distance transform (DT) [35] image is predicted. During the training phase, this image is created from the mask of the manually segmented image and used for U-Net training. Each cell distance map is normalised to have a maximum value of one. In the inference phase, DT image is predicted with a trained network from the input image and used for instance cell segmentation (Fig. 1(b)). A robust maxima detector (described above) is then applied for seed generation, and these are used together with the predicted DT image and the predicted foreground image as an input to the above-described MCW algorithm.

BP – In the second approach, the individual cell masks are converted into cell boundary pixels (BP), which can be obtained by dilatation of individual cells and determining their overlap (see Fig. 1(b)). The amount of this dilatation was selected by manual tuning of a disc structuring element with a radius equal to 8 pixels. In post-processing, the eroded foreground (with an optimised size structuring element) is divided by the predicted boundary pixels. The resulting seeds are filtered with an area filter and minimal distance between centroids. Besides these seeds, the negative of the original QPI image is used as the input image for the watershed.

BE – In the third approach, besides the foreground/background, a binary eroded (BE) foreground is predicted, where erosion will make a larger separation between individual cells (see Fig. 1(b)). The amount of erosion was selected by manual tuning of a disc structuring element with a radius equal to 4 pixels. A separation of not completely separated cells is done by the post-processing with DT , robust maxima detector (without threshold T) and watershed algorithm. The combination of DT with the watershed algorithm is a standard approach for splitting the connected objects in their narrowest connection (i.e., it splits the shapes in the narrowest points). With a robust maxima detector, this separation is regularised with minimal centroid distance and h -maxima transform. Output is used as a seed for the second watershed algorithm, where the negative of the original QPI image is used as the input image for MCW.

NDT – In the fourth approach, the neighbour distance transform (NDT) [14] is applied, and this predicted parametric map instead of the normalised DT image is used. An image transformed by NDT contains the values of distance to the closest cell (see Fig. 1(b)), and background pixels are set to zero. This can be obtained with multiple DT calculations – for each cell, we can calculate DT from other cells and use the region inside this cell for NDT . In the post-processing

phase, the predicted *NDT* image is multiplied by the eroded foreground (eroded with a circular structuring element of optimised size), and a robust maxima detector is applied to obtain seeds. Finally, *NDT* image is used as an input image for MCW.

2.3. Self-supervised pretraining

Self-supervised pretraining has proved to be an efficient method to improve the efficiency of different tasks in machine learning. We have tested four different approaches and their combination. All these approaches distort the input image in a different way, and the U-net is learned to restore these distortions using MSE loss. The principle of pretraining is shown in Fig. 2(a), and few examples of distorted images are shown in Fig. 2(c). The following distortions were applied:

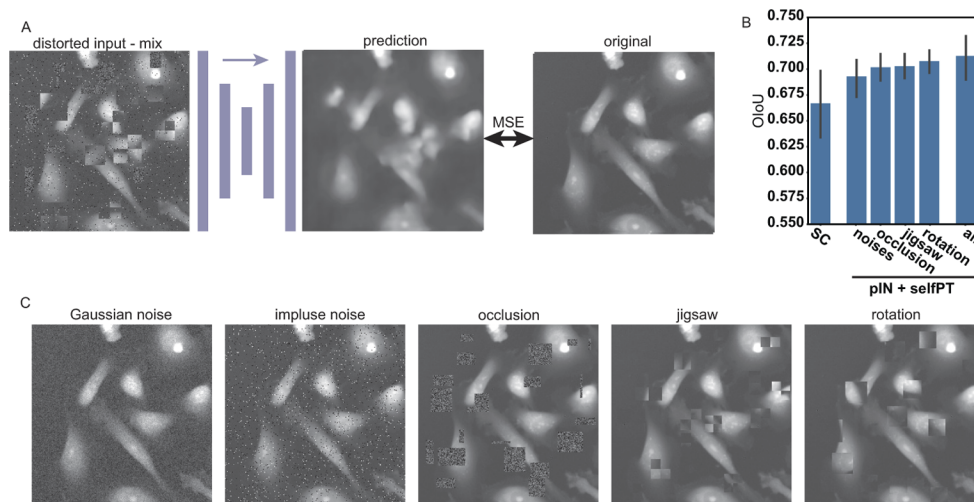


Fig. 2. Principle and results of pretraining methods: (a) Schematic example of principle of pretraining with example of network input and output – distorted input image created by combination of several augmentation techniques should be restored by U-Net; mean squared error (MSE) between restored and original undistorted image is used for pretraining. (b) Results of comparison of proposed mixed pretraining (*all*) with individual image distortions used for pretraining and without any pretraining (*SC*); Distance transform (*DT*) method is used for all evaluations; for pretraining methods the network pretrained also on ImageNet beforehand (*pIN+selfPT*) is used; results are for 5-fold validation and bar plots show average and standard deviation, (c) Example images of individual distortion methods.

Additive noise with two different distributions – impulse and Gaussian noise, where the probability of each pixel being corrupted and the standard deviation were optimised, respectively. For impulse noise, only some pixels were corrupted (with a specified probability), but with large maximal noise values of 5-times the average image standard deviation.

Occlusion with rectangular blocks – we have generated a set of randomly sized rectangular blocks at random positions and replaced them with a Gaussian noise with a standard deviation equal to the average standard deviation. The number of blocks and the maximal block length of the rectangle edge were optimised. This, also known as image inpainting, is another straightforward method for pretraining of segmentation networks [26].

Rotation of square block – Another pretraining method of classification network is the prediction of rotation [24], which is also adapted in our application. Specifically, for rotation, we have rotated random square blocks, where the number of blocks and block size were optimised.

Reordering of four quadrants – Jigsaw puzzle pretraining [25] was also adapted for the segmentation network. It was implemented as a selection of random square blocks, which were split into four quadrants, and these quadrants were reshuffled. Similarly to the rotation and occlusion, the number of blocks and block size were optimised.

For the final self-supervised pretraining, a mixture of all these distortions was used, where the parameters of individual distortions were optimised using Bayesian optimisation. [32], but only a single validation fold was used during optimisation. For the optimised parameter ranges and optimal values see [Supplement 1](#).

2.4. Evaluation metrics

For all results, the values of 5-fold validation are presented, where for each fold a new random train/validation/test split was applied. For the evaluation of semantic segmentation, the results can be easily evaluated with binary Intersection over Union (IoU):

$$IoU = \frac{R \cap S}{R \cup S} = \frac{TP}{TP + FP + FN} \quad (1)$$

where R is the set of cell pixels of Ground Truth (GT) mask, S is the set of cell pixels in the algorithm result of semantic segmentation. TP , FP , and FN are the number of true positive, false positive, and false negative pixels, respectively. Similarly, F_1 score (Dice coefficient) can be used, where it can be converted to IoU with monotonic transformation (maintaining the algorithm ranking). IoU for binary (semantic) segmentation will be denoted as BIoU.

For the evaluation of instance segmentation, an object-based metric is required. The object score F_1 is defined in [13], such that the segmented cell is considered as a true positive if its IoU with the corresponding cell in the GT mask is higher than the selected threshold. A similar metric called Average Precision (AP) applies object-wise IoU instead of F_1 score: $AP_T = TP/(TP + FP + FN)$, where subscript T denotes the threshold for the object to be considered as TP [17]. Thus, AP can be calculated for various thresholds with a minimum threshold of 0.5 to ensure the uniqueness of the assignment of GT cells to the resulting cell.

However, AP does not produce a single number, which would be easier to handle and would be more suitable for optimisation tasks. On the other hand, SEG score (used in the cell tracking challenge [36]) combines pixel segmentation accuracy with the correctness of identification of individual cells in a single number. In SEG, for every GT object, the segmented object with the largest IoU is found. If IoU for any GT object is smaller than 0.5, then IoU for this object is set to zero. Next, the average IoU of all GT objects is calculated. Again, the threshold 0.5 ensures that each GT object can be paired with only one segmented object. As SEG score does not contain false positive (FP) objects, we are using a more strict modification, where FP objects are counted as additional zero values for the calculation of this metric, and it will be denoted as Object-wise IoU (OIoU).

3. Experimental results

Several experimental setups were conducted to compare and test the proposed approach. First of all, U-Net based approaches were compared to non-DL approaches. Then, for the best method (i.e. post-processing pipeline), we show the improvement achieved by the self-supervised pretraining. Afterwards, we evaluated the *non-DL transferability* of the whole pipeline to different cell types by re-optimisation of the post-processing parameters only, without pretraining the DL network – non-DL-transfer.

3.1. Deep-learning and classical methods comparison

Comparisons of the proposed U-Net based approaches and non-DL approaches are shown on Fig. 3(a)-(c). For non-DL approaches, implementations from [4] and [11] were used. Specifically,

simple threshold segmentation combined with the fast radial symmetry transform detection, $sST + dFRST$, and simple threshold segmentation combined with distance transform-based detection ($sST + dDT$) from [4] was used; improved iterative thresholding (IIT) from [11] and implementation of iterative thresholding (denoted as *Loewke*) [10] was used. Similarly, the parameters of non-DL methods were optimised on the validation set.

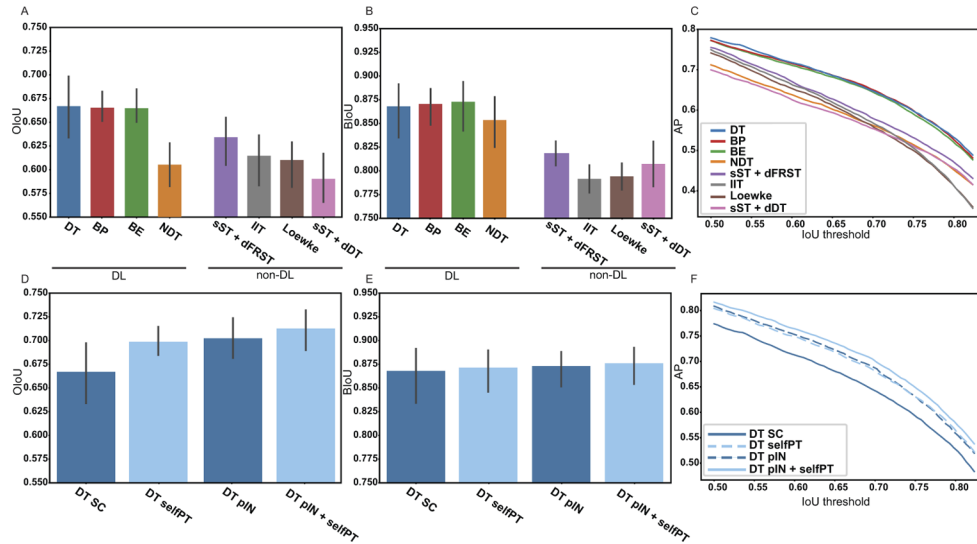


Fig. 3. Results of the proposed methods and available non-DL methods for QPI cell segmentation. Upper row: comparison of three proposed DL approaches with existing non-DL methods is shown using different metrics: (a) – Object-wise Intersection over Union (OIoU), (b) – Binary Intersection over Union (BIOU), and (c) – Average Precision (AP) dependent on IoU thresholds. Lower row: comparison of proposed network trained from scratch (*SC*), proposed self-supervised pretraining (*selfPT*), ImageNet pretrained network (*pIN*), and ImageNet pretrained network with additional proposed self-supervised pretraining (*pIN+selfPT*) is shown in (d) – OIoU, (e) – BIOU, and (f) – AP dependent on IoU thresholds, where all results are shown for distance transform (*DT*) method. Results are for 5-fold validation, where AP shows average value and bar plots show average and standard deviation.

U-Net based approaches except for *NDT* method performed very similarly in all metrics (OIoU, BIOU and AP) with 0.667, 0.665 and 0.664 of OIoU for *DT*, *BP* and *BE* methods, respectively. Proposed *NDT* approach performed significantly worse, similarly to non-DL methods with OIoU 0.606, 0.634, 0.615, 0.610, 0.590, for *NDT*, $sST + dFRST$, *IIT*, *Loewke* and $sST + dDT$, respectively. They are also significantly worse in the other metrics, BIOU and AP.

Examples of results for PC-3 and PNT1A cells are shown in Fig. 4. For less densely clustered cells (i.e., easier to segment), all methods performed similarly with relatively good results. However, for densely clustered cells, noticeable differences between methods can be observed. The following evaluations are performed for the best performing *DT* method in OIoU metric, which was chosen as the main optimisation metric in this paper.

3.2. Self-supervised pretraining evaluation

Self-supervised pretraining (denoted as *selfPT*) using the optimised setting (optimised by Bayesian optimisation on the validation set) was compared to the network trained from scratch (denoted as *SC*) and the network pretrained on ImageNet dataset (denoted as *pIN*); moreover, we have tried to use ImageNet pretrained network and retrain it again using the proposed self-supervised pretraining (denoted as *pIN+selfPT*). As shown in Fig. 3(d)-(f), *selfPT* and *pIN* networks performed similarly

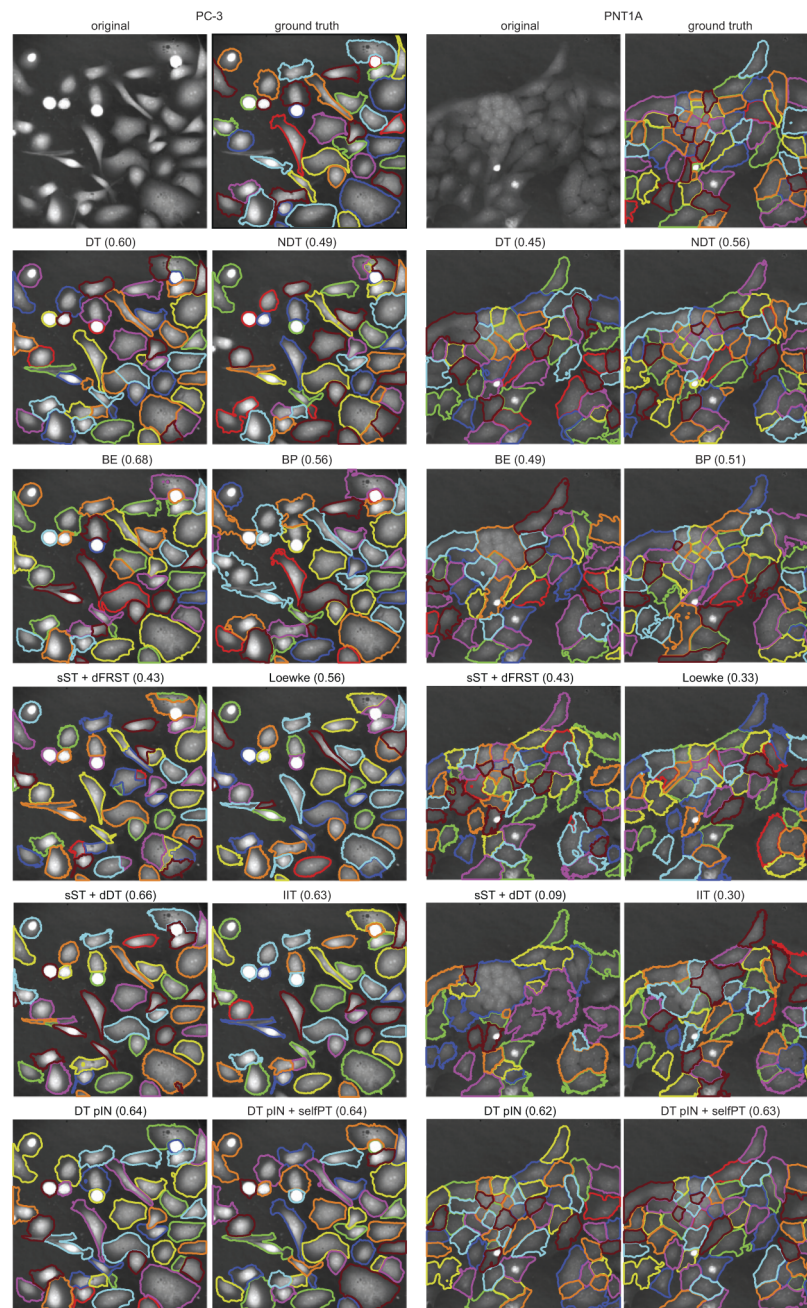


Fig. 4. Example of results for one field of view of PNT1A and PC3 cells using different segmentation methods. The numbers in brackets represent OIoU for individual images in this example. Colour contours represent individual cell borders.

with OIoU 0.702 and 0.699, respectively; however, its combination ($pIN+selfPT$) leads to an additional improvement to 0.712 OIoU. The same trend is also kept for other metrics – BIoU and AP. The examples in Fig. 4 shows that $pIN+selfPT$ significantly improved the segmentation of densely clustered cells.

Comparisons of the proposed mixed pretraining techniques with only individual distortion pretraining are shown in Fig. 2(b). The best result was achieved by the proposed optimised mixture of all distortions (OIoU 0.712). The OIoU values for individual distortions were 0.693, 0.702, 0.703 and 0.708 for noise, occlusion, jigsaw, and rotation, respectively. The rotation performs best of the individual methods, and noise performs worst.

Moreover, the dependence of SC , pIN and $pIN + selfPT$ networks on the amount of training data is shown in Fig. 5, where you can see OIoU and BIoU performance of these networks for 10%, 25%, 50% and 100% of randomly selected training data. It shows that $pIN + selfPT$ keeps a reasonable performance of 0.55 OIoU even in very low data regimes – 10% of training data, while SC and pIN networks failed with significantly lower values of 0.24 OIoU and 0.31 OIoU, respectively.

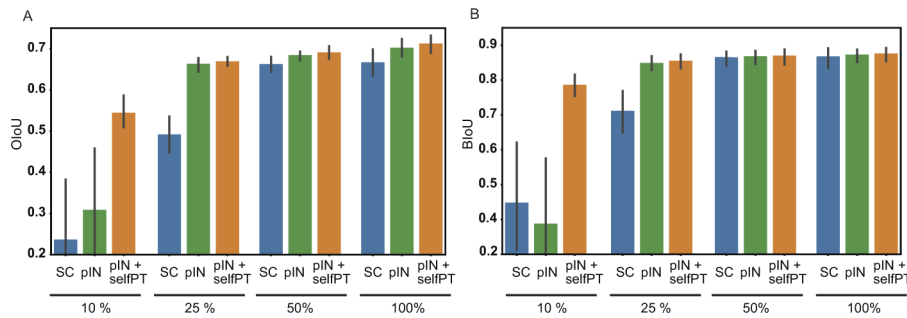


Fig. 5. Results of the networks with the reduced training set to 10%, 25%, 50% and 100% of randomly selected training data. Mean results of Object-wise Intersection over Union (OIoU) and Binary Intersection over Union (BIoU) are shown on (a) and (b), respectively. SC , pIN and $pIN + selfPT$ are results for training from scratch, ImageNet pretraining and proposed self-supervised pretraining with network pretrained on ImageNet beforehand, respectively. Numbers are averages of 5-fold validation and lines in bar plots represent standard deviations.

3.3. Transferability analysis between PC-3 and PNT1A

The proposed pipeline consists of deep neural network training and post-processing pipeline parameters optimisation, which opens the possibility of transfer to different cell lines just by adjustment of a few post-processing parameters. For this, a combination of training/optimisation/testing on PC-3 or PNT1A only, and on the mix of both cell lines (PC-3+PNT1A) was evaluated, and the results are shown in Fig. 6. Moreover, the results of other pretraining strategies (SC , pIN and $pIN+selfPT$) are presented, which shows the effect of pretraining on the *non-DL transferability*.

Segmentation results of PC-3 cells for the $pIN+selfPT$ reached a very similar value for the network trained and optimised on PC-3 – 0.71 OIoU (see Fig. 6(c)) and for the network trained on PNT1A and optimised on PC-3 – 0.69 OIoU. Even the $pIN+selfPT$ network trained and optimised on PNT1A performed well on PC-3 cells – 0.69 OIoU. However, the SC network trained and optimised on PNT1A performed significantly worse on PC-3 cells – 0.62 OIoU.

For PNT1A cells, there are larger differences between the network trained on PNT1A and the network transferred on PNT1A by post-processing pipeline parameters optimisation. The $pIN+selfPT$ network trained and optimised on PC-3 reached on PNT1A cells OIoU 0.61. When it was *non-DL transferred* by post-processing parameters optimisation (i.e., trained on PC-3

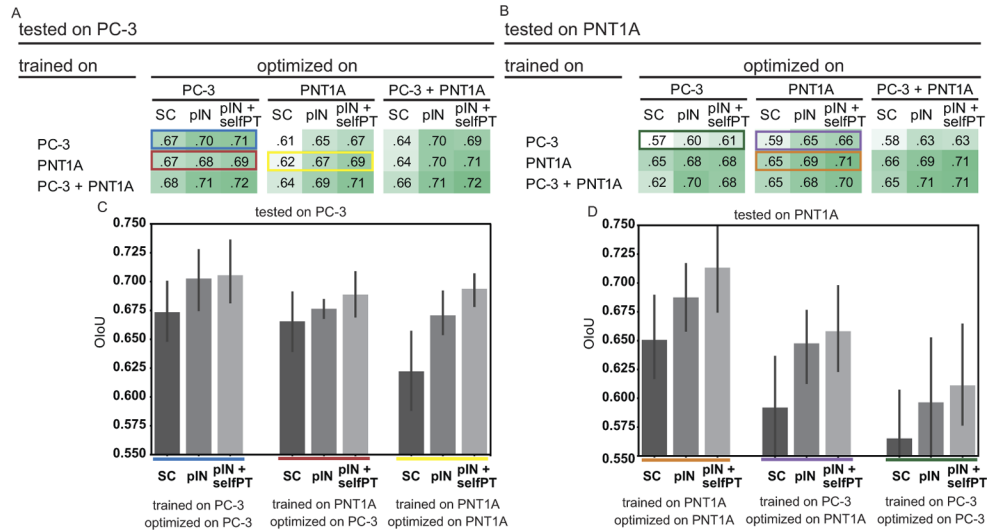


Fig. 6. Results of transferability between PC-3 and PNT1A by optimisation of post-processing parameters. Tables (a-b) show combinations of cell type used for training of the deep network, cell type used for optimisation of post-processing pipeline parameters, and cell type used for evaluation (PC-3 – images with PC-3 cell line only, images with PNT1A – PNT1A cell line only, PC-3 + PNT1A – mix of images containing both above mentioned cell lines). Important selected values from tables are shown on plots (c-d) using the corresponding colours. *SC*, *pIN* and *pIN + selfPT* are results for training from scratch, ImageNet pretraining and proposed self-supervised pretraining with network pretrained on ImageNet beforehand, respectively. Numbers are averages of 5-fold validation and lines in bar plots represent standard deviations.

and optimised on PNT1A) it significantly improved the performance to 0.66. In comparison, the network trained and optimised using PNT1A reached 0.71. For the *SC* network trained and the *pIN* network, the same trend is kept – the OIoU performance is lower than our proposed pretraining scheme.

3.4. Non-DL transfer application to unseen cell lines

In the last experiment we used different cell types (PNT1A or PC-3) to examine the ability of *non-DL* transferability to diverse cell lines. For this purpose, we prepared a new manually labelled dataset (15 images for validation and 10 images for testing) for three diverse cell lines from pretraining data – G361, HOB, and A2058.

For each new cell type, we took the model originally trained on a mixture of PC-3 and PNT1A images and the parameters of post-processing pipeline were optimised using 15 validation images of the new cell type (G361/HOB/A2058). The achieved results for individual cell types are presented in Fig. 7 for the three pretraining strategies – *SC*, *pIN* and *pIN + selfPT*. It can be seen that both self-supervised pretraining and *non-DL transfer* significantly improve the OIoU value. The highest influence of *non-DL transfer* is achieved on the A2058 cell type, because these cells are morphologically distinctly different from the cells used for training (see Supplement 1, Fig. S2), while for morphologically similar G361 cells the influence is much lower. For *pIN + selfPT* network and HOB cells, non-DL transfer leads to a slight decrease of mean OIoU performance. However, this is a small difference on HOB cells, which are similar to PC3/PNT1A cells, thus, there is no room for improvement with *non-DL transfer*.

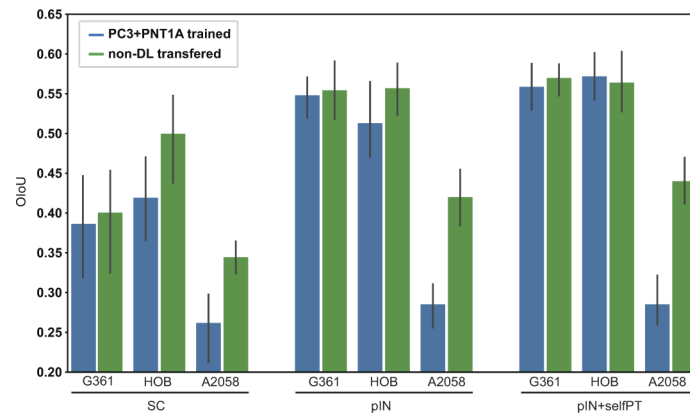


Fig. 7. Results of non-Deep-Learning (*non-DL*) transferability to new cell types by optimisation of post-processing parameters. Results of network trained and optimised on mixture of PC3 and PNT1A with original post-processing parameters are shown in blue. Results of network trained on mix of PC-3 and PNT1A and post-processing parameters optimised on the individual new cell types are shown in blue. *SC*, *pIN* and *pIN + selfPT* are results for training from scratch, ImageNet pretraining and proposed self-supervised pretraining with network pretrained on ImageNet beforehand, respectively. G361, HOB and A2058 are cell lines used for evaluation. Numbers are averages of Object-wise Intersection over Union (OIoU) of 5-fold validation and lines in bar plots represent standard deviations.

4. Discussion

The results in this paper have shown the superiority of DL methods for QPI cell segmentation over the classical approaches. However, the gap between DL and non-DL approaches is not as significant as in other applications. As our results show, the network performance gradually increases with the amount of training data. The main advantage of the deep learning approach might only be evident on orders of magnitude larger datasets. However, the proposed dataset is relatively large and adding new manually segmented cells is always connected to the limited precision of a human observer. Furthermore, it must be noted that QPI is typically easily segmentable, and thus, the application of non-DL methods might provide satisfactory results, particularly for adherent cells with lower density. A benefit of DL approach arises in more difficult tasks (segmentation of high cell density with complex shapes).

In our approach, we proposed and compared four post-processing pipelines (DT, BP, BE, NDT). We observed that the prediction of *NDT* image performed by the U-net has very low quality. Thus, we conclude that this representation is difficult to predict with the selected network, and for this reason it shows the worst performance of proposed DL methods. Proposed *BE* and *BP* methods performed similarly as *DT*; however, *BE* and *BP* methods use the original QPI image in its pipeline for creation of the final cell separation line. Therefore, they are modality-dependent and cannot be directly used for e.g., differential interference contrast microscopy images [1]. The selected *DT* approach uses only the estimated *DT* image, which can be, in principle, predicted from *any* imaging modality.

In addition, a new evaluation metric named OIoU for the evaluation of instance segmentation is also proposed, which summarises the correctness of detection of individual objects together with the precision of their segmentation. Compared to AP, it produces just a single number, which can be used for the optimisation of the method; and compared to SEG score, it also penalises false positive detections.

Different implementations of DL approach in this paper were performed with two particular networks – one for foreground prediction and the second for the prediction of the image with separated cells. This approach ensures avoiding ‘fight over features’ for these two tasks; however, the training of a single network for the prediction of both images together is, in principle, possible with the benefit of the faster inference.

Self-supervised pretraining (*selfPT*) on unlabelled data has shown similar performance as ImageNet pretrained network (*pIN*); however, self-supervised pretraining of ImageNet pretrained network together (*pIN+selfPT*) further increased the performance. Moreover, *pIN+selfPT* performed significantly better especially with a very small amount of training data, which can be useful in practical applications. Furthermore, we also performed a simple experiment with the influence of an amount of pretraining data on the result and we showed that pretraining improves the results even with a small amount of pretraining data, and does not improve with additional data significantly (see [Supplement 1](#), Fig. S1). Different distortions applied during the pretraining have used the same framework of restoration of the distorted image, which enables to include another type of distortion. However, the combination of single distortion has led to only a negligible result improvement in comparison to patch rotation. Thus, further investigation of self-supervised pretraining may bring new findings. The self-supervised pretraining also may not be the best approach, how to efficiently utilise unlabelled data. For example, a multi-task network with self-supervised tasks and a segmentation task trained synchronously may perform even better [37]. We leave those investigations to future work.

The main advantage of optimisation in the post-processing phase is the number of optimised parameters (only four parameters in our implementation). This enables easy application of the already trained DL network to a slightly different task (i.e., different cell lines with different morphological properties) and might be a part of the solution leading to green AI strategies [38]. Moreover, these parameters can also be adjusted manually without optimisation. The success of this approach will depend on the dissimilarity of the particular tasks. However, in our case, we have shown that *non-DL transfer* is an efficient way to adjust the whole pipeline from PC-3 cell to PNT1A cell segmentation and vice versa. Similarly, we confirmed that both *non-DL transfer* and self-supervised pretraining improved the network performance on distinct cell types, which were not used for network training. The combination with self-supervised pretraining provides an efficient way to achieve higher segmentation precision without the need for large labelled data for the new task.

5. Conclusion

In this paper, the U-Net-based method for robust adherent cell segmentation for quantitative phase microscopy image was designed and optimised. Four different U-Net based methods for instance cell segmentation were tested, and three of these methods achieved very similar results. These DL-based methods outperformed several well-performing non-DL methods. However, the gap between DL and non-DL methods is not so significant on a dataset of this size. Additionally, a novel self-supervised pretraining method based on image reconstruction from multiple distortions was proposed, where the proposed mixture of distortions achieved better results than each individual distortion. This improved the segmentation performance from 0.67 to 0.70 of Object-wise IoU, compared to a network trained from scratch. Another important characteristic of the proposed approach is the post-processing pipeline with adjustable parameters. This concept enables to test the non-deep learning transferability between different cell types without retraining the DL model just by optimisation of a few parameters of these post-processing pipelines. A manually segmented dataset for QPI cell segmentation (449 images) is published simultaneously with this paper (*QPI_Seg_PNT1A_PC3*) with additional unlabelled data (1,819 images) for self-supervised pretraining (*QPI_Cell_unlabelled*).

Funding. Grantová Agentura České Republiky (18-24089S).

Acknowledgments. Computational resources were supplied by 'e-Infrastruktura CZ' (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovation Infrastructures. Jaromir Gumulec was supported by funds from the Faculty of Medicine, Masaryk University for junior researchers.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Ref. [39] (manually labelled *QPI_Seg_PNT1A_PC3* dataset and unlabelled *QPI_Cell_unlabelled* for self-supervised pretraining), and the code is available in Ref. [40].

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. T. Vicar, J. Balvan, J. Jaros, F. Jug, R. Kolar, M. Masarik, and J. Gumulec, "Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison," *BMC Bioinf.* **20**(1), 360–425 (2019).
2. M. K. Kim, "Principles and techniques of digital holographic microscopy," *SPIE Reviews* **1**, 1–51 (2010).
3. D. Roitshtain, L. Wolbromsky, E. Bal, H. Greenspan, L. L. Satterwhite, and N. T. Shaked, "Quantitative phase microscopy spatial signatures of cancer cells," *Cytom. Part A* **91**(5), 482–493 (2017).
4. T. Vicar, M. Raudenska, J. Gumulec, and J. Balvan, "The quantitative-phase dynamics of apoptosis and lytic cell death," *Sci. Rep.* **10**(1), 1566–1612 (2020).
5. T. Blasi, H. Hennig, H. D. Summers, F. J. Theis, J. Cerveira, J. O. Patterson, D. Davies, A. Filby, A. E. Carpenter, and P. Rees, "Label-free cell cycle analysis for high-throughput imaging flow cytometry," *Nat. Commun.* **7**(1), 10256 (2016).
6. L. Kastl, M. Isbach, D. Dirksen, J. Schnekenburger, and B. Kemper, "Quantitative phase imaging for cell culture quality control," *Cytom. Part A* **91**(5), 470–481 (2017).
7. H. Alanazi, A. J. Canul, A. Garman, J. Quimby, and A. E. Vasdekis, "Robust microbial cell segmentation by optical-phase thresholding with minimal processing requirements," *Cytom. Part A* **91**(5), 443–449 (2017).
8. V. L. Calin, M. Mihailescu, N. Tarba, A. M. Sandu, E. Scarlat, M. G. Moisesescu, and T. Savopol, "Digital holographic microscopy evaluation of dynamic cell response to electroporation," *Biomed. Opt. Express* **12**(4), 2519–2530 (2021).
9. B. Kemper, H. Eilers, T. Klein, K. Brinker, and S. Ketelhut, "Quantitative phase imaging-based machine learning approaches for the analysis of adherent and suspended cells," *Proc. SPIE* **11649**, 116490B (2021).
10. N. O. Loewke, S. Pai, C. Cordeiro, D. Black, B. L. King, C. H. Contag, B. Chen, T. M. Baer, and O. Solgaard, "Automated cell segmentation for quantitative phase microscopy," *IEEE Transactions on Med. Imaging* **37**(4), 929–940 (2018).
11. T. Vicar, J. Chmelik, and R. Kolar, "Cell segmentation in quantitative phase images with improved iterative thresholding method," in *8th European Medical and Biological Engineering Conference*, (Springer International Publishing, 2021), pp. 233–239.
12. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention 2015 (MICCAI 2015)*, vol. 9351 of *Lecture Notes in Computer Science*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, eds. (Springer, 2015), pp. 234–241.
13. J. C. Caicedo, J. Roth, A. Goodman, T. Becker, K. W. Karhohs, M. Broisin, C. Molnar, C. McQuin, S. Singh, F. J. Theis, and A. E. Carpenter, "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images," *Cytom. Part A* **95**(9), 952–965 (2019).
14. F. Lux and P. Matula, "Cell segmentation by combining marker-controlled watershed and deep learning," <https://arxiv.org/abs/2004.01607>.
15. T. Scherr, K. Löffler, M. Böhlend, and R. Mikut, "Cell segmentation and tracking using distance transform predictions and movement estimation with graph-based matching," <https://arxiv.org/abs/2004.01486> (2020).
16. U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, "Cell detection with star-convex polygons," in *Medical Image Computing and Computer-Assisted Intervention 2018*, vol. 11071 of *Lecture Notes in Computer Science*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer, 2018), pp. 265–273.
17. C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nat. Methods* **18**(1), 100–106 (2021).
18. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2017), pp. 2961–2969.
19. K. Eder, T. Kutscher, A. Marzi, Álvaro Barroso, J. Schnekenburger, and B. Kemper, "Automated detection of macrophages in quantitative phase images by deep learning using a mask region-based convolutional neural network," *Proc. SPIE* **11655**, 54 (2021).
20. Y.-H. Lin, K. Y.-K. Liao, and K.-B. Sung, "Automatic detection and characterization of quantitative phase images of thalassemic red blood cells using a mask region-based convolutional neural network," *J. Biomed. Opt.* **25**(11), 1–14 (2020).
21. F. Yi, I. Moon, and B. Javidi, "Automated red blood cells extraction from holographic images using fully convolutional neural networks," *Biomed. Opt. Express* **8**(10), 4466–4479 (2017).
22. Z. Zhang, K. W. Leong, K. V. Vliet, G. Barbastathis, and A. Rivasio, "Deep learning for label-free nuclei detection from implicit phase information of mesenchymal stem cells," *Biomed. Opt. Express* **12**(3), 1683–1706 (2021).

23. T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," <https://arxiv.org/abs/2006.10029>.
24. S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," <https://arxiv.org/abs/1803.07728>.
25. C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2015), pp. 1422–1430.
26. D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 2536–2544.
27. M. Prakash, T.-O. Buchholz, M. Lalit, P. Tomancak, F. Jug, and A. Krull, "Leveraging self-supervised denoising for image segmentation," in *Proceedings of IEEE 17th International Symposium on Biomedical Imaging* (IEEE, 2020), pp. 428–432.
28. Z. Zeng, Y. Xulei, Y. Qiyun, Y. Meng, and Z. Le, "Sese-net: Self-supervised deep learning for segmentation," *Pattern Recognit. Lett.* **128**, 23–29 (2019).
29. T. Slaby, P. Kolman, Z. Dostal, M. Antos, M. Lostak, and R. Chmelik, "Off-axis setup taking full advantage of incoherent illumination in coherence-controlled holographic microscope," *Opt. Express* **21**(12), 14747–14762 (2013).
30. M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning* (PMLR, 2019), pp. 6105–6114.
31. F. Meyer, "Topographic distance and watershed lines," *Signal Proc.* **38**(1), 113–125 (1994).
32. J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems* (2012), pp. 2951–2959.
33. K. Thirusittampalam, M. J. Hossain, O. Ghita, and P. F. Whelan, "A novel framework for cellular tracking and mitosis detection in dense phase contrast microscopy images," *IEEE J. Biomed. Heal. Informatics* **17**(3), 642–653 (2013).
34. F. Nogueira, "Bayesian optimization: open source constrained global optimization tool for python," Github, 2014, <https://github.com/fmfn/BayesianOptimization>.
35. C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Transactions on Pattern Analysis Mach. Intell.* **25**(2), 265–270 (2003).
36. V. Ulman, M. Maška, K. E. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic, I. Smal, K. Rohr, J. Jalden, H. Blau, O. Dzyubachyk, B. Lelieveldt, P. Xiao, Y. Li, S.-Y. Cho, D. A. C. J.-C. Olivo-Marin, C. C. Reyes-Aldasoro, J. A. Solis-Lemus, R. Bensch, T. Brox, J. Stegmaier, R. Mikut, S. Wolf, F. A. Hamprecht, T. Esteves, P. Quelhas, Ö. Demirel, L. Malmström, F. Jug, P. Tomancak, E. Meijering, A. Muñoz-Barrutia, M. Kozubek, and C. Ortiz-de Solorzano, "An objective comparison of cell-tracking algorithms," *Nat. Methods* **14**(12), 1141–1152 (2017).
37. S. Reiß, C. Seibold, A. Freytag, E. Rodner, and R. Stiefelhagen, "Every annotation counts: multi-label deep supervision for medical image segmentation," in *Conference on Computer Vision and Pattern Recognition* (2021), pp. 9532–9542.
38. R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Commun. ACM* **63**(12), 54–63 (2020).
39. T. Vicar, J. Chmelik, R. Jakubicek, L. Chmelikova, J. Gumulec, J. Balvan, I. Provaznik, and R. Kolar, "Annotated quantitative phase microscopy cell dataset of various adherent cell lines for segmentation purposes," Zenodo, 2021, <https://doi.org/10.5281/zenodo.5153251>.
40. T. Vicar, J. Chmelik, R. Jakubicek, L. Chmelikova, J. Gumulec, J. Balvan, I. Provaznik, and R. Kolar, "Deep-qpi-cell-segmentation: self-supervised pretraining for transferable quantitative phase image cell segmentation," Github, 2021, <https://github.com/tomasvicar/Deep-QPI-Cell-Segmentation>.